View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repositorio Institucional de la Universidad de Alicante

Procesamiento del Lenguaje Natural, Revista nº 53, septiembre de 2014, pp 151-154 recibido 12-04-14 revisado 02-06-14 aceptado 02-06-14

ATTOS: Análisis de Tendencias y Temáticas a través de Opiniones y Sentimientos

ATTOS: Trend Analysis and Thematic through Opinions and Sentiments

L. Alfonso Ureña López*, Rafael Muñoz Guillena, José A. Troyano Jiménez***, Mª Teresa Martín Valdivia***
*SINAI - Universidad de Jaén. Campus Las Lagunillas s/n, 23071, Jaén
**GPLSI - Universidad de Alicante. San Vicente del Raspeig, s/n, 03690, Alicante
***ITALICA - Universidad de Sevilla. Av. Reina Mercedes, s/n. 41012 - Sevilla
laurena@ujaen.es, rafael@dlsi.ua.es, troyano@us.es, maite@ujaen.es

Resumen: El proyecto ATTOS centra su actividad en el estudio y desarrollo de técnicas de análisis de opiniones, enfocado a proporcionar toda la información necesaria para que una empresa o una institución pueda tomar decisiones estratégicas en función a la imagen que la sociedad tiene sobre esa empresa, producto o servicio. El objetivo último del proyecto es la interpretación automática de estas opiniones, posibilitando así su posterior explotación. Para ello se estudian parámetros tales como la intensidad de la opinión, ubicación geográfica y perfil de usuario, entre otros factores, para facilitar la toma de decisiones. El objetivo general del proyecto se centra en el estudio, desarrollo y experimentación de técnicas, recursos y sistemas basados en Tecnologías del Lenguaje Humano (TLH), para conformar una plataforma de monitorización de la Web 2.0 que genere información sobre tendencias de opinión relacionadas con un tema.
Palabras clave: Análisis de Opiniones y Sentimientos, Tecnologías del Lenguaje Humano, Recuperación de Información, Clasificación de Opiniones, Procesamiento de Lenguaje Natural

Abstract: The ATTOS project will be focused on the study and development of Sentiment Analysis techniques. Thanks to such techniques and resources, companies, but also institutions will be better understood which is the public opinion on them and thus will be able to develop their strategies according to their purposes. The final aim of the project is the automatic interpretation of such opinions according to different variables: opinion, intensity, geographical area, user profile, to support the decision process. The main objective of the project is the study, development and evaluation of techniques, resources and systems based on Human Language Technologies to build up a monitoring platform of the Web 2.0 that generates information on opinion trends related with a topic.
Keywords: Opinion and Sentiments Analysis, Human Language Technology, Information Retrieval, Opinion Classification, Natural Language Processing

1 Introducción

La interacción actual de los usuarios de la Sociedad de la Información es muy participativa. Los usuarios expresan sus puntos de vista, opiniones que llegan de forma inmediata al resto de usuarios a través de la Web 2.0 (foros, blogs, microblogs, redes sociales, etc.). Como consecuencia, la cantidad de información de carácter subjetivo y lenguaje informal se ha multiplicado en los últimos años, surgiendo así nuevos retos para las Tecnologías del Lenguaje Humano (TLH), como son el tratamiento de distintos registros de uso con diferentes grados de informalidad, el estudio de distintas actitudes subjetivas o el multilingüismo.

Las actuales herramientas para las TLH no son directamente aplicables a estos nuevos usos y medios de comunicación o son, simplemente, inadecuadas, por lo que se hace esencial adaptar

ISSN 1135-5948
© 2014 Sociedad Española para el Procesamiento del Lenguaje Natural

y crear nuevos recursos, métodos y herramientas para su tratamiento.

El objetivo global del proyecto es, por tanto, el estudio, desarrollo y experimentación de recursos, diferentes técnicas y sistemas basados en TLH para el desarrollo y la comprensión de las expresiones subjetivas y el lenguaje informal en diversos dominios de aplicación, así como el desarrollo de una plataforma online de monitorización de diversos tipos de objetos de acuerdo a la información subjetiva e informal extraída de varias fuentes online de información. En este nuevo escenario, los sistemas deben incorporar recursos, herramientas y sistemas que descubrirán la subjetividad de la información en todos sus contextos (espacial, temporal y emocional) analizando la dimensión multilingüe, y su aplicación en diversos dominios.

2 Objetivos

Los objetivos generales del proyecto ATTOS son[1]:

- Creación, adaptación y mejora de recursos, técnicas y herramientas que modelan el lenguaje subjetivo e informal generado en diversas fuentes de información (blogs, microblogs, redes sociales, etc.). Tratamiento del lenguaje emocional, la multilingualidad y la aplicación a entornos concretos.
- Desarrollo de subsistemas inteligentes de procesamiento (recuperación, tratamiento, comprensión y descubrimiento) de la información adaptados a las nuevas formas de comunicación con capacidad de interpretar y valorar el contexto del mensaje.
- Integración de los recursos, herramientas y sistemas desarrollados para el análisis de la subjetividad en una plataforma web de monitorización, cuya validez se demostrará sobre varios escenarios de uso concreto de distintos ámbitos (turismo, política, empresarial, comercio electrónico, etc.). Evaluación de la plataforma. Promoción de las líneas de investigación del proyecto mediante la participación y organización de actividades en campañas, congresos, talleres, seminarios y redes temáticas

Para la consecución del objetivo global y el desarrollo óptimo de las diferentes líneas de actuación del proyecto, se propuso la coordinación de tres subproyectos complementarios cuyos objetivos específicos abarcan los objetivos globales planteados, y cuya reunificación proporcionará el valor añadido que se buscaba en la coordinación.

El subproyecto ATTOS -Análisis de Tendencias y Temáticas a través de Opiniones y Sentimientos- que lleva a cabo el equipo de investigación de la Universidad de Jaén, tiene como objetivo central la construcción de una plataforma de procesamiento inteligente que integre las técnicas desarrolladas por todos los equipos de este proyecto para la explotación de la información subjetiva y que sea fácilmente adaptable a diversos dominios de aplicación. Así el subproyecto SOTTA -Semantic Opinion Techniques for Tendencies Analysis- que lleva a cabo el equipo de la Universidad de Alicante, tiene como objetivo principal el desarrollo de una herramienta de análisis de tendencias en función a perfiles de usuarios, incorporando técnicas que permitan identificar y resolver la presencia de metáforas, ironía y sarcasmo en textos subjetivos. Y finalmente el subproyecto ACOGEUS -Análisis de COntenidos GEnerados por USuarios- a cargo del grupo de la Universidad de Sevilla, cuyo objetivo es la identificación de fuentes online con información subjetiva y recuperación de dicha información, creando recursos propios para los dominios a abordar, así como el desarrollo de técnicas que permitan identificar diversos registros del lenguaje (ofensivo, violento, etc.).

3 Propuesta

El objetivo general del proyecto se centra en el estudio, desarrollo y experimentación de diferentes técnicas y sistemas basados en Tecnologías del Lenguaje Humano (TLH) para el desarrollo de una plataforma de tratamiento de información subjetiva y lenguaje informal, afrontando los actuales retos de la comunicación digital. En este nuevo escenario, los sistemas deben incorporar capacidades de razonamiento que descubrirán la subjetividad de la información desde diversas dimensiones: multilingüe, espacial, temporal y emocional.

[1] Sitio web: http://attos.ujaen.es

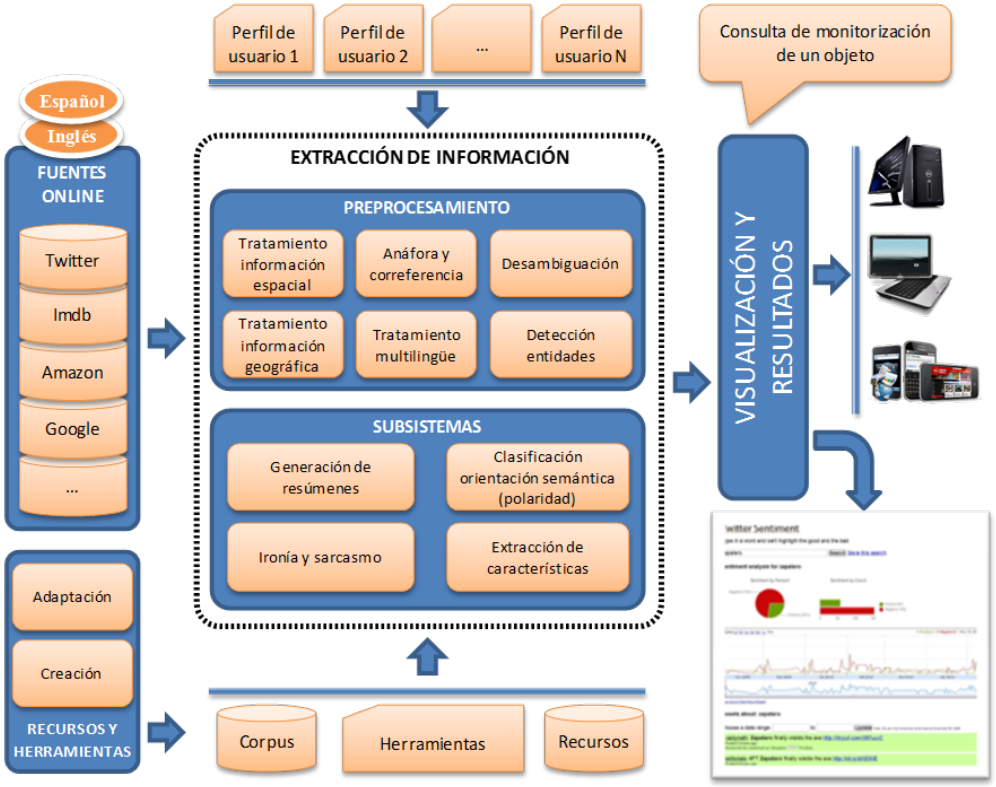

Figura 1: Arquitectura general del sistema

La figura 1 muestra la manera en la que se pueden integrar distintos componentes para construir un sistema capaz de procesar distintas fuentes online y extraer indicadores de utilidad mediante la aplicación de distintas tecnologías del lenguaje humano.

El diseño de los módulos del plan de trabajo propuesto se corresponde con las líneas de actuación marcadas en los objetivos del proyecto.

- En el módulo 1 se gestiona el proyecto y se diseñan mecanismos de coordinación que permitan una comunicación fluida y una colaboración eficiente entre los distintos miembros del proyecto.
- El módulo 2 se centra en el desarrollo y adaptación de recursos, herramientas y métodos de TLH para el modelado, análisis y tratamiento de información subjetiva e informal.
- En el módulo 3 se desarrollan los sistemas de detección y tratamiento de la información subjetiva y su tratamiento, su especialización en diversos dominios de aplicación y el desarrollo de una plataforma online de visualización y presentación de resultados.
- El módulo 4 contempla las actividades necesarias para la evaluación de la utilidad de la plataforma tanto interna como externa, así como la promoción, coordinación y participación en diferentes foros de evaluación.
- Finalmente, mediante el módulo 5, se creará un plan estratégico para diseminar los resultados tanto científicamente como mediáticamente para lograr la mayor difusión posible y facilitar la transferencia de tecnología a la empresa.

Respecto al enfoque científico, este proyecto supone un reto en el modo de abordar nuevos registros del lenguaje, como es la información digital subjetiva y el lenguaje informal. El problema actual es afrontar el tratamiento de una creciente cantidad de información en los nuevos registros que la Web 2.0 contiene: información textual en formatos muy variados expresada en muchas ocasiones de manera espontánea sin la precisión, formalidad ni corrección de los textos normativos. Desde la perspectiva computacional, requiere un replanteamiento de los métodos y técnicas de adquisición automática de conocimiento para tratar nuevas unidades y características, además de aquellas que son tradicionalmente aceptadas.

4 Resultados

En el tiempo en el que el proyecto lleva en ejecución, los trabajos realizados se han

materializado en diferentes contribuciones como publicaciones en revistas, congresos, organización de eventos o participación en evaluaciones competitivas. En esta sección comentaremos brevemente algunos de estos resultados que constituyen una muestra significativa de los avances que se están consiguiendo en el proyecto.

En al ámbito de las redes sociales, y en concreto en Twitter, en (Cotelo et al., 2014) se definió un método para obtener de forma automática consultas adaptativas a partir de un conjunto de *hashtags* semilla. Este método es especialmente interesante para poder capturar tweets relacionados con una temática, contemplando de forma automática los términos que puedan ir apareciendo en el transcurso de los diálogos colaborativos que permite esta red social.

En el contexto del análisis de opiniones, en (Cruz et al., 2013) se presenta un sistema de extracción adaptable a dominio que permite identificar opiniones en textos escritos por usuarios extrayendo la característica sobre la que se opina (p.e. el precio) y la valoración correspondiente (positiva o negativa). También se desarrolló un algoritmo de polaridad en 6 niveles mediante la modificación de un algoritmo de ranking (RA-SR) mediante la utilización de bigramas un puntuador de skipgrams (Fernández et al, 2013). En (Molina-González et al., 2013) se presenta una lista de palabras indicadoras de opinión en español de dominio general, así como una metodología para la adaptación de lexicones de palabras de opinión a un dominio concreto. También se han obtenido unos primeros resultados en la clasificación de la polaridad en redes sociales. En (Montejo-Ráez et al., 2013) se presenta un sistema de clasificación de la polaridad sobre *tweets*, cuya mayor aportación es el método de desambiguación utilizado, el cual utiliza información del contexto para mejorar la exactitud de la desambiguación, y la inclusión de términos relacionados para el cálculo de la polaridad del *tweet*.

En la detección de la subjetividad se desarrolló un método a nivel de oraciones basado en la desambiguación subjetiva del sentido de las palabras. Para ello se extiende un método de desambiguación semántica basado en agrupamiento de sentidos para determinar cuándo las palabras dentro de la oración están siendo utilizadas de forma subjetiva u objetiva (Ortega et al. 2013).

Agradecimientos

El proyecto ATTOS está financiado por el Ministerio de Economía y Competitividad con número de referencia TIN2012-38536-C03-01, TIN2012-38536-C03-02 y TIN2012-38536-C03-03. Con el apoyo de la Red Temática TIMM: Tratamiento de Información Multimodal y Multilingüe. (TIN2011-13070-E).

Bibliografía

Cotelo, J.M., Cruz, F.L. Troyano, J.A. 2014. Dynamic topic-related tweet retrieval. *JASIST*. 65(3): 513-523.

Cruz, F.L., Troyano, J.A., Enríquez, F., Ortega, F.J., Vallejo, C.G. 2013. 'Long autonomy or long delay?' The importance of domain in opinion mining. *Expert Syst. Appl*. 40(8): 3174-3184.

Fernández, J., Gómez, J.M.; Martínez, P., Montoyo, A, Muñoz, R. 2013 Sentiment Analysis of Spanish Tweets Using a Ranking Algorithm and Skipgrams. TASS 2013: Taller de Análisis de Sentimientos en la SEPLN / Workshop on Sentiment Analysis at SEPLN Madrid, Spain, SEPLN.

Molina-González, M. Dolores, Martínez-Cámara, Eugenio, Martín-Valdivia, M. Teresa, Perea-Ortega, Jose M. 2013. Semantic Orientation for Polarity Classification in Spanish Reviews. *Expert Systems with Applications*. 40(18):7250-7257.

Montejo-Ráez, Arturo, Martínez-Cámara, Eugenio, Martín-Valdivia, M. Teresa, Ureña-López, L. Alfonso. 2014. A Knowledge-Based Approach for Polarity Classification in Twitter. *JASIST*. 65(2):414-425.

Ortega, R.; Fonseca, A.; Gutierrez, Y.; Montoyo, A.2013 Improving Subjectivity Detection using Unsupervised Subjectivity Word Sense Disambiguation. Revista Procesamiento del Lenguaje Natural, 51.